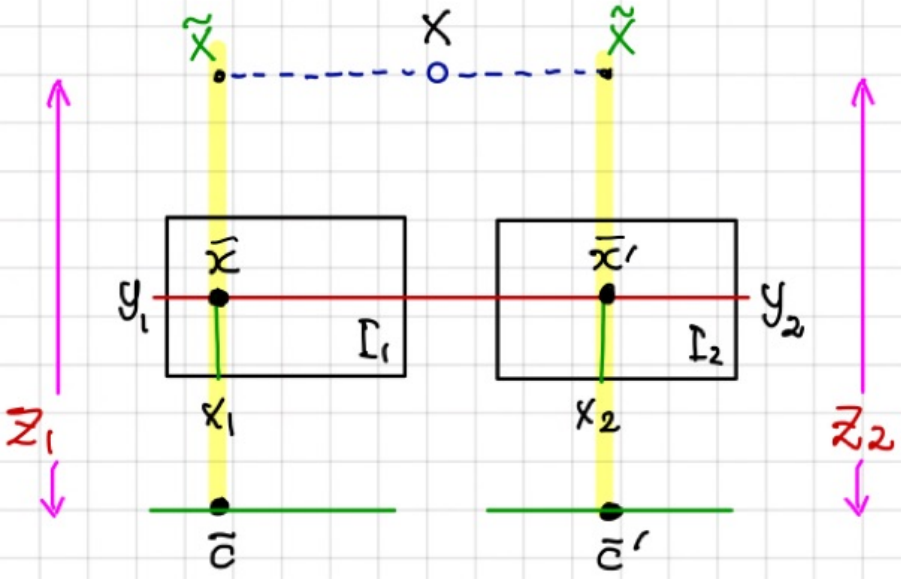
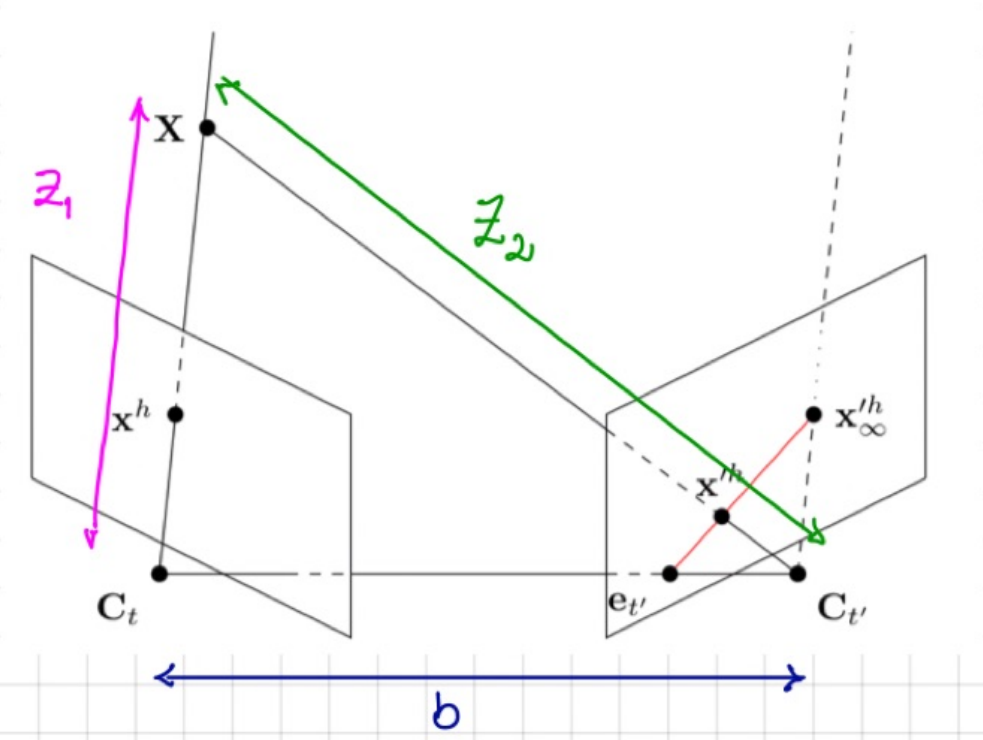
LECTURE 8 : DEPTH FROM VIDEO

For a pair of rectified images, there is only one disparity or depth map:



Though, before being rectified, there are 2 depth maps:



Q: How to compute disparity of non-rectified images?

A: Two ways:

(1) Rectify the images → Epipolar rectification

(2) Using epipolar geometry

⇓

Epipolar-geometry implicit rectification is preferable, since less steps are required in the computation.

↓

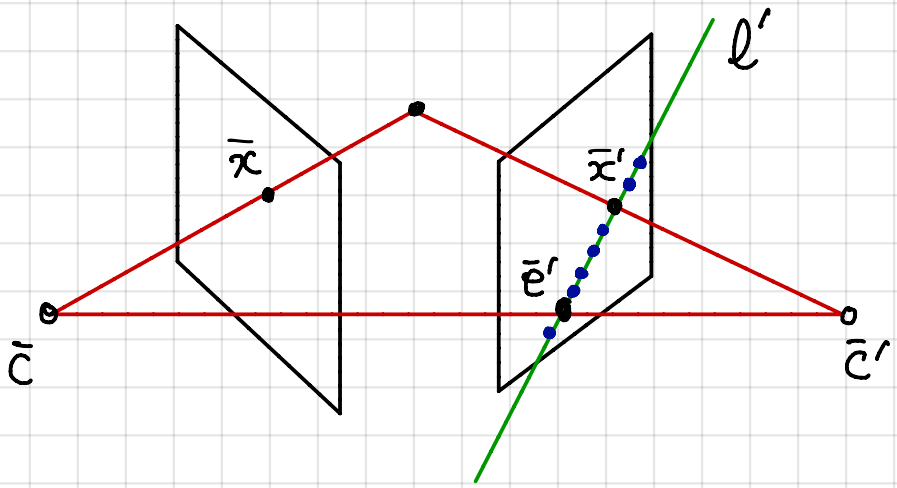
The basic formulation:

$d = \frac{1}{Z}$; d = disparity, Z = depth

$$\bar{x}' \sim \mathbb{K}'\mathbb{R}'\mathbb{R}^T\mathbb{K}^{-1}\bar{x} + d\,\mathbb{K}'\mathbb{R}'\left(\bar{C} - \bar{C}'\right)$$

Given $\bar{x}$, $\mathbb{P} = \mathbb{K}[\mathbb{R}|\bar{t}]$ and $\mathbb{P}' = \mathbb{K}'[\mathbb{R}'|\bar{t}']$, we want to find d using the formula, such that $I(\bar{x}) = I'(\bar{x}')$.

The formula can generate all points on the green line below, when we change the value of d:

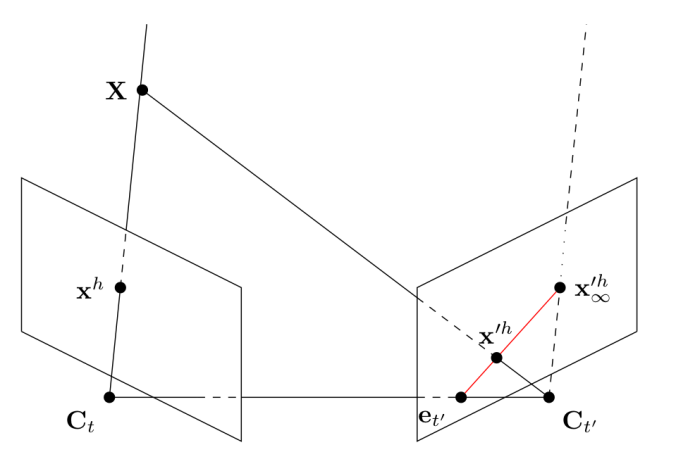


[4] Derivation of Epipolar Rectification

To understand how we can get this equation:

$$\bar{x}' \sim \mathbb{K}'\mathbb{R}'\mathbb{R}^T\mathbb{K}^{-1}\bar{x} + d\,\mathbb{K}'\mathbb{R}'\left(\bar{C} - \bar{C}'\right).$$

consider:



$$\bar{x}^h = \mathbb{P}X = \mathbb{K}[\mathbb{R}\,|\,\bar{t}]X$$

Assume the existence of X_∞ along the line that passes through X, where X_∞ is located at infinity:

$$X_\infty = \begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{X}_\infty \\ 0 \end{bmatrix}$$

$$\underset{3\times1}{\bar{x}^h} = \underset{3\times3}{\mathbb{K}}\underset{3\times3}{[\mathbb{R}\,|\,\bar{t}]}\underset{4\times1}{X_\infty} = \underset{3\times3}{\mathbb{K}}\underbrace{[\mathbb{R}\,|\,\bar{t}]}_{3\times4}\underset{4\times1}{\begin{bmatrix} \hat{X}_\infty \\ 0 \end{bmatrix}} = \underset{3\times3}{\mathbb{K}}\underset{3\times3}{\mathbb{R}}\underset{3\times1}{\hat{X}_\infty}$$

Thus:

$$\hat{X}_\infty = \mathbb{R}^T\mathbb{K}^{-1}\bar{x}^h$$

Projecting X_∞ onto the right image:

$$\bar{x}'^h_\infty = \mathbb{P}'X_\infty = \mathbb{K}'[\mathbb{R}'\,|\,\bar{t}']\begin{bmatrix} \hat{X}_\infty \\ 0 \end{bmatrix} = \mathbb{K}'\mathbb{R}'\hat{X}_\infty$$

$$= \mathbb{K}'\mathbb{R}'\mathbb{R}^T\mathbb{K}^{-1}\bar{x}^h$$

What we want to find is $\bar{x}'^h$, and not $\bar{x}'^h_\infty$.

Basic idea of finding $\bar{x}'^h$:

To find $\bar{x}'^h$, we can start the search from $\bar{x}'^h_\infty$. Then from $\bar{x}'^h_\infty$, we can trace along the epipolar line in the direction of the epipolar point, $\bar{e}_{t'}$:

$$\underset{3\times1}{\bar{e}_{t'}} = \underset{3\times4}{P'}\ \underset{4\times1}{\bar{C}_t} = K'[R'|\bar{t}']\,\bar{C}_t$$

$$= K'R'\left[\underset{3\times3}{I}\ \middle|\ \underset{3\times1}{-[R']^{-1}\bar{t}'}\right]\underset{4\times1}{\bar{C}_t}$$

Recall: $\bar{C}_{t'} = -(R')^{-1}\bar{t}'$

$$= \underset{3\times3}{K'}\ \underset{3\times3}{R'}\left[\underset{3\times1}{\bar{C}_t} - \underset{3\times1}{\bar{C}_{t'}}\right]$$

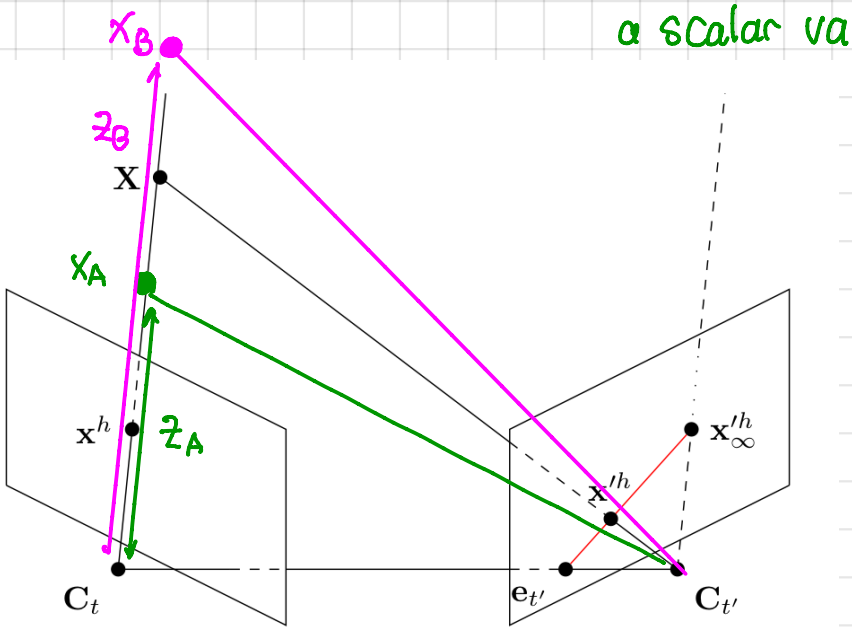
Therefore:

$$\underset{3\times1}{\bar{x}'^h} \sim \underset{3\times1}{\bar{x}'^h_\infty} + d\ \underset{3\times1}{\bar{e}_{t'}} \quad \rightarrow \text{the line parametric equation}$$

$$\bar{x}'^h \sim K'R'R^TK^{-1}\bar{x}^h + d\,\bar{e}_{t'}$$

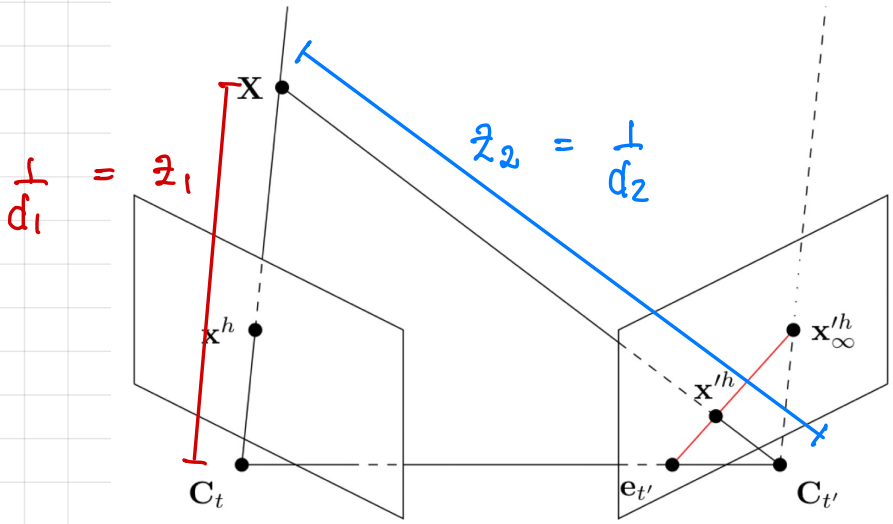
$$\underset{3\times1}{\bar{x}'^h} \sim \underset{3\times3}{K'}\ \underset{3\times3}{R'}\ \underset{3\times3}{R^T}\ \underset{3\times3}{K^{-1}}\ \underset{3\times1}{\bar{x}^h} + d\ \underset{3\times3}{K'}\ \underset{3\times3}{R'}\left[\underset{3\times1}{\bar{C}_t} - \underset{3\times1}{\bar{C}_{t'}}\right]$$

d: a scalar value



The shorter the depth (Z_A), the larger the value of d, vice versa.

This also indicates d determines the depth of point X.



d = disparity
z = depth

In the figure, with respect to X, there are 2 depths:

- $z_1 = 1/d_1$ is the depth from C_t to X
- $z_2 = 1/d_2$ is the depth from $C_{t'}$ to X

$$\underset{3\times1}{\bar{x}'^h} \sim \underset{3\times1}{\bar{x}'^h_\infty} + d_1 \underset{3\times1}{\bar{e}_{t'}}$$

→ You need to normalize x'^h so that: $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} x/z \\ y/z \\ 1 \end{pmatrix}$

Q: How to justify that this equation is correct?

A: Consider the following cases:

If $d = 0$, meaning $z = \infty$:

$$\bar{x}'^h = \bar{x}'^h_\infty$$

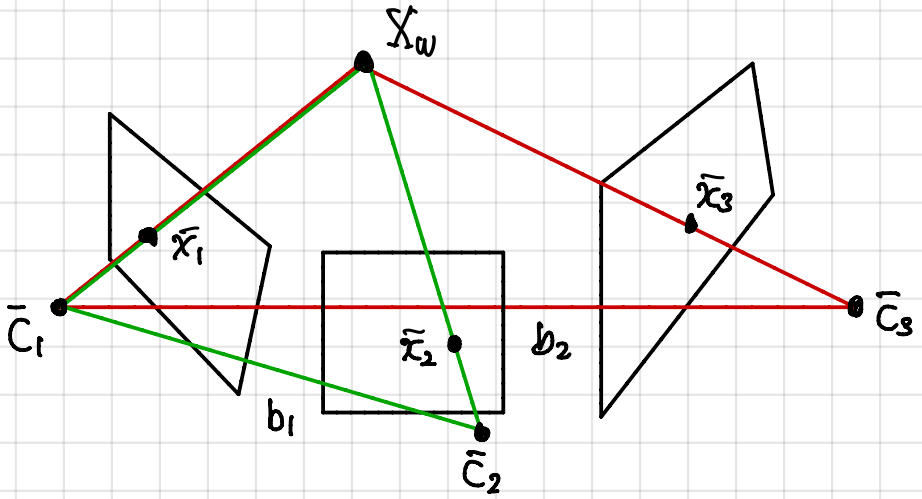
If $d = \infty$, meaning $z = 0$ (the 3D point is at $\bar{C}_t$):

$$\bar{x}'^h = \bar{x}'^h_\infty + \infty\, \bar{e}' = \bar{e}'$$

} see the illustration above

If $0 < d < \infty$, then $\bar{x}'^h$ must lie on the epipolar line!

[5] Epipolar Rectification for Multiple Images



[o] For image 1 & 2 :

$$\bar{x}_2 = \bar{x}_{2\infty} + d_{12}\,\bar{e}_2 = K_2 R_2 R_1^T K_1^{-1} \bar{x}_1 + d_{12} K_2 R_2 (\bar{C}_1 - \bar{C}_2)$$

[o] For image 1 & 3 :

$$\bar{x}_3 = \bar{x}_{3\infty} + d_{13}\,\bar{e}_3 = K_3 R_3 R_1^T K_1^{-1} \bar{x}_1 + d_{13} K_3 R_3 (\bar{C}_1 - \bar{C}_3)$$

$$\Downarrow$$

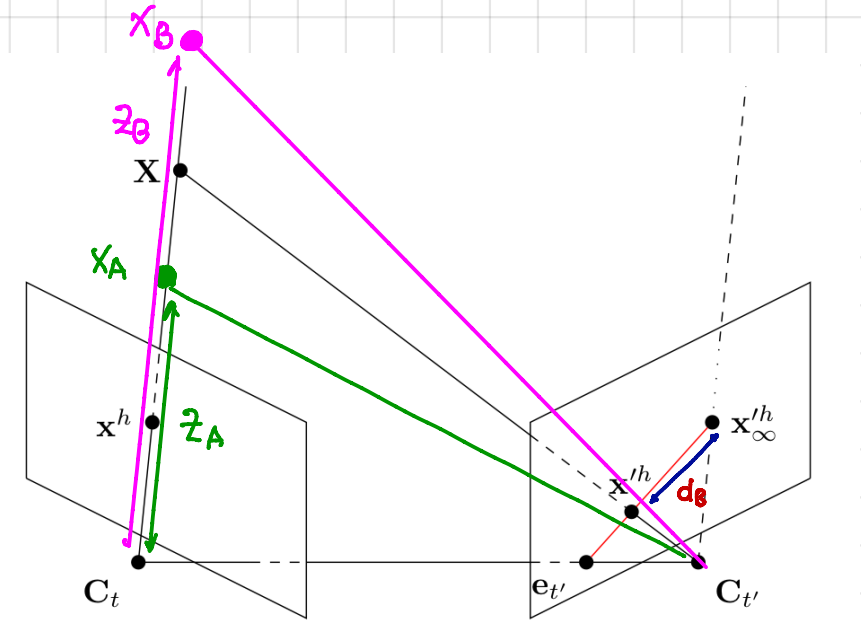
$$\boxed{d_{12} = d_{13}}$$

Implying : $I_2(\bar{x}_2) = I_1(\bar{x}_1)$

$I_3(\bar{x}_3) = I_1(\bar{x}_1)$

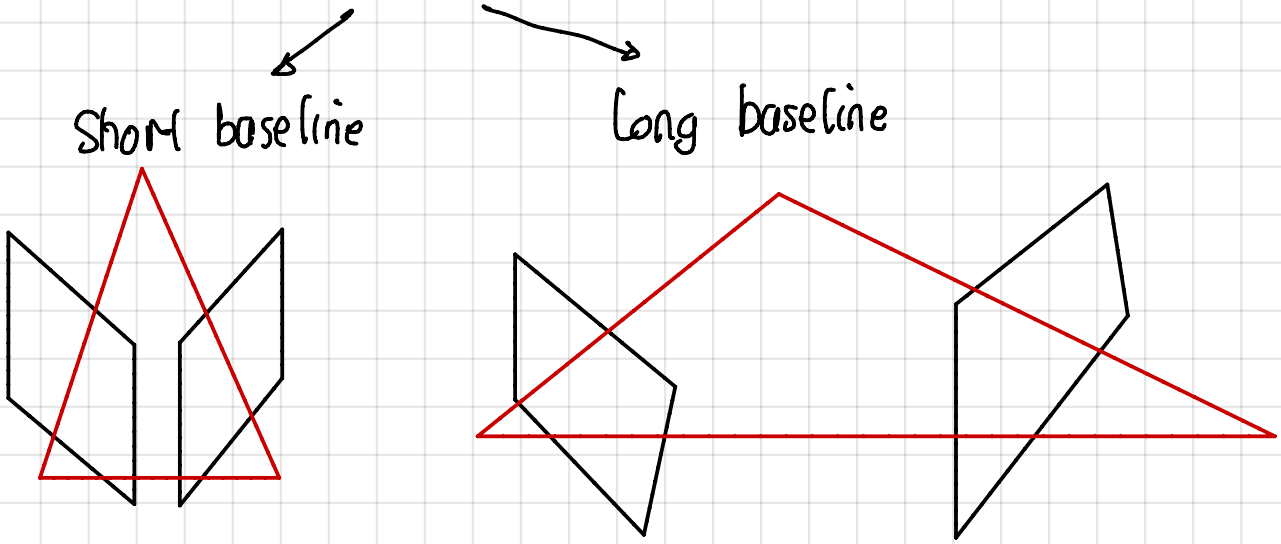
Q : Why $d_{12} = d_{13}$?

A : Because d_{12} & d_{13} indicate the same depth of X from C_1.



[6] Multiple Baseline Stereo = Depth from Video

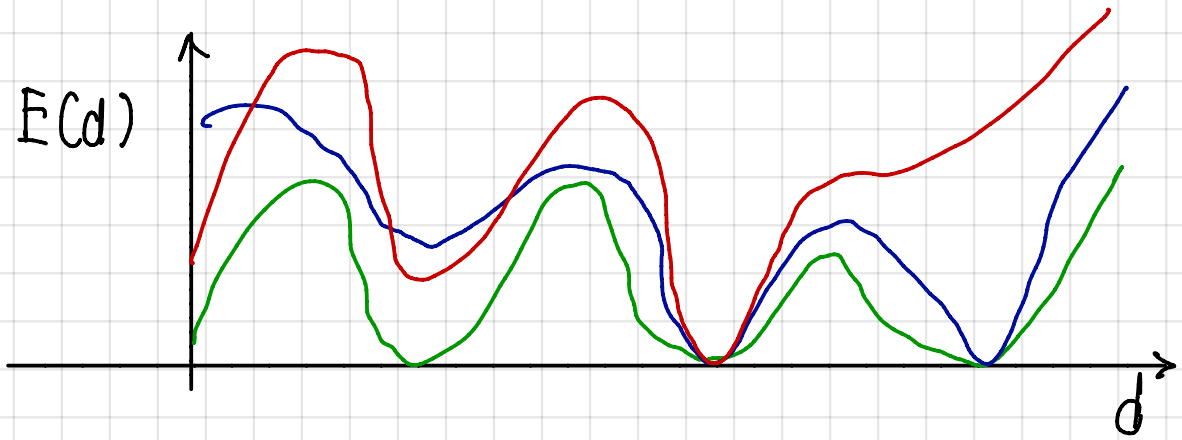
Disparity values depends on the baseline



For estimating disparity, a longer baseline is generally better, since b in the equation acts like a magnifier. However, longer baselines tend to suffer from occlusions.

↓

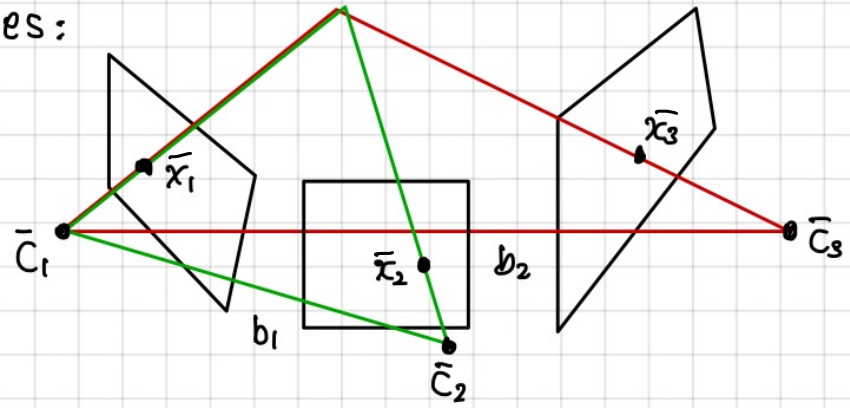
Multiple baselines allow us to get the benefits of short and long baselines:



The green line indicates that there are many candidates of d that make $E(d)=0$ ⇒ ambiguity!

[7] Depth from Video (Multiple Baselines)

Multiple baselines:



Two main steps

(1) Disparity Initialization

(2) Bundle Optimization

Purpose: To have an initial depth map of each frame.

Input: $\{I_t, P_t\}_{t=1}^{N}$; $t = 1 \dots N$

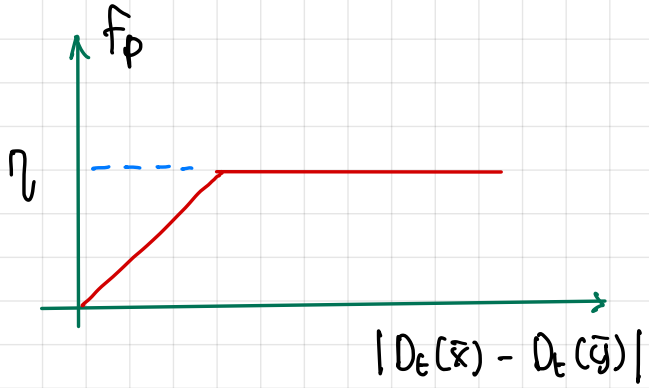
Output: $\{D_t\}_{t=1}^{N}$ (the disparity maps)

N: the number of frames.

MRF:

$$D_t^{init} = \underset{\{d_{min} \dots d_{max}\}}{\operatorname{argmin}} \sum_{\bar{x}} \left[(1 - U(\bar{x})) f_d^{init}(\bar{x}, D_t(\bar{x})) + \sum_{\bar{y} \in N_{\bar{x}}} \lambda(\bar{x}, \bar{y}) f_p(D_t(\bar{x}), D_t(\bar{y})) \right]$$

- Prior term: $f_p(D_t(\bar{x}), D_t(\bar{y})) = \min\left(|D_t(\bar{x}) - D_t(\bar{y})|, \eta \right)$



To prevent f_p from being too large, since the difference between two disparity values shouldn't be that large.

Main reference: "Consistent depth map recovery from a video sequence." TPAMI, 2009.

MRF:

$$D_t^{init} = \underset{\{d_{min} \dots d_{max}\}}{\operatorname{argmin}} \sum_{\bar{x}} \left[1 - U(\bar{x}) f_d^{init}(\bar{x}, D_t(\bar{x})) + \sum_{\bar{y} \in N_{\bar{x}}} \lambda(\bar{x}, \bar{y}) f_p(D_t(\bar{x}), D_t(\bar{y})) \right]$$

- Data term:

Use all other frames (prev. & subsequent frames)

$$f_d^{init}(\bar{x}, D_t(\bar{x})) = \sum_{t'}^{N} \rho_c(\bar{x}, D_t(\bar{x}), I_t, I_{t'})$$

e.g.: $t = 1$, $t' = 2, 3, \dots, N$

where:

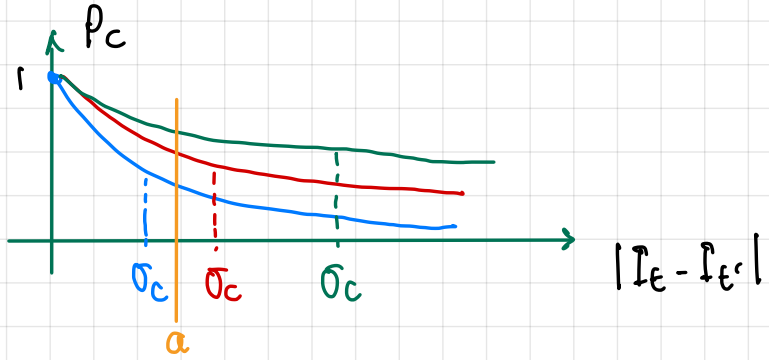
$$\rho_c(\bar{x}, d, I_t, I_{t'}) = \frac{\sigma_c}{\sigma_c + |I_t(\bar{x}) - I_{t'}(l_{t,t'}(\bar{x}, d))|}$$

photo consistency

$l_{t,t'}(\bar{x}_t, d)$ means:

$$\underset{3\times1}{\bar{x}_{t'}} = \underset{3\times3}{K_{t'}} \underset{3\times3}{R_{t'}} \underset{3\times3}{R_t^T} \underset{3\times3}{K_t^{-1}} \underset{3\times1}{\bar{x}_t} + \underset{1\times1}{d} \underset{3\times3}{K_{t'}} \underset{3\times3}{R_{t'}} \underset{3\times1}{[\bar{C}_t - \bar{C}_{t'}]}$$

σ_c implies a variable to control the tolerance of the photo consistency:



if $|I_t - I_{t'}| = a$, then different values of σ_c will produce different ρ_c. The smaller σ_c (blue line) penalize $|I_t - I_{t'}|$ more. Meaning, even though $|I_t - I_{t'}|$ is relatively small, ρ_c will be smaller compared with the other values of σ_c (red and green lines).

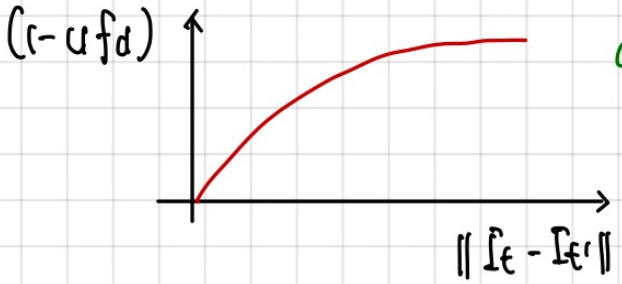
MRF :

$$D_t^{init} = \underset{\{d_{min} \dots d_{max}\}}{\operatorname{argmin}} \sum_{\bar{x}} \left[(1 - U(\bar{x})\, f_d^{init}(\bar{x}, D_t(\bar{x})) + \sum_{\bar{y} \in N_{\bar{x}}} \lambda(\bar{x}, \bar{y})\, f_p(D_t(\bar{x}), D_t(\bar{y})) \right]$$

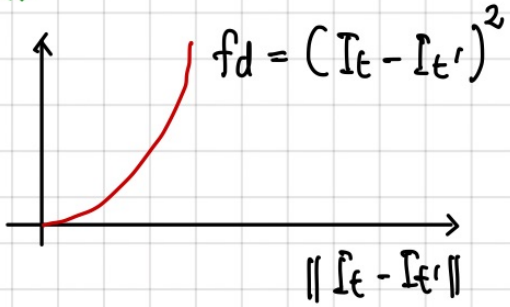
$$U(\bar{x}) = \frac{1}{\max_{D_t(\bar{x})} f_d(\bar{x}, D(t))}$$

→ the normalization factor. To ensure the values ranging from 0 to 1.

The graph of the overall data term:



more robust than our prev. cost function →



If there is noise $\| I_t - I_{t'} \|$ will be large, hence $(I_t - I_{t'})^2$ will be also large, though the actual cost shouldn't be that large. Using $(1-uf_d)$ will reduce the effect of noise in the optimization.

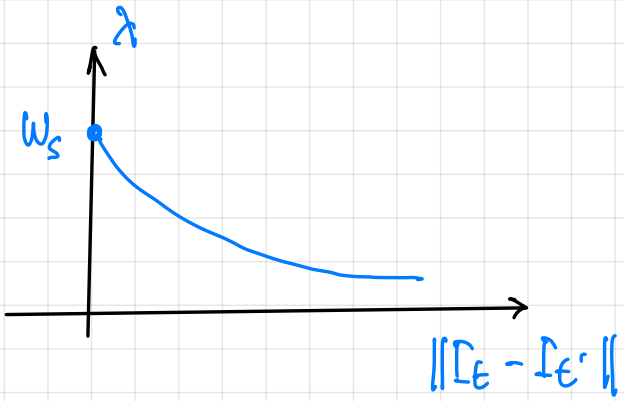
- Weighting factor λ :

MRF :

$$D_t^{init} = \underset{\{d_{min} \dots d_{max}\}}{\operatorname{argmin}} \sum_{\bar{x}} \left[1 - U(\bar{x}) f_d^{init}(\bar{x}, D_t(\bar{x})) + \sum_{\bar{y} \in N_{\bar{x}}} \lambda(\bar{x}, \bar{y}) f_p(D_t(\bar{x}), D_t(\bar{y})) \right]$$

Goal : To preserve the discontinuities. λ is defined to encourage the disparity discontinuities to be consistent with intensity / color discontunities.

$$\lambda(\bar{x}, \bar{y}) = w_s \frac{U_\lambda(\bar{x})}{\| I_t(\bar{x}) - I_t(\bar{y}) \| + \varepsilon}$$



where :

$$U_\lambda(x) = \frac{|N_x|}{\sum_{\bar{y}' N_{\bar{x}}} \frac{1}{\| I_t(\bar{x}) - I_t(\bar{y}') \| + \varepsilon}}$$

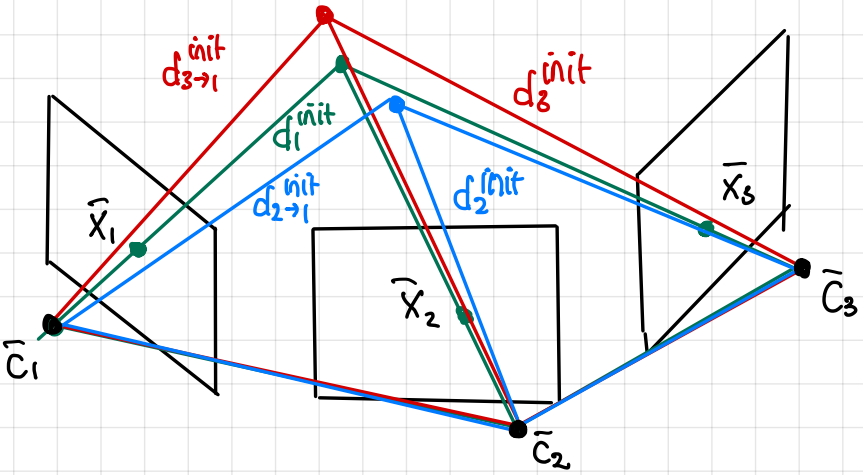
w_s = the smoothness strength

$$\lambda(\bar{x}, \bar{y}) = w_s \frac{|N_x|}{(\| I_t(\bar{x}) - I_t(\bar{y}) \| + \varepsilon) \sum_{\bar{y}' \in N_{\bar{x}}} \frac{1}{\| I_t(\bar{x}) - I_t(\bar{y}') \| + \varepsilon}}$$

1. The smaller the intensity / color difference ($\| I_t(\bar{x}) - I_t(\bar{y}) \|$) between two neighboring pixels, the larger λ is. A larger λ will encourage the smoothness constraint more.
2. U_λ is to normalize $\| I_t(\bar{x}) - I_t(\bar{y}) \|$ so that λ should be within a certain range depending on w_s.

[8] Bundle Optimization

Problem :

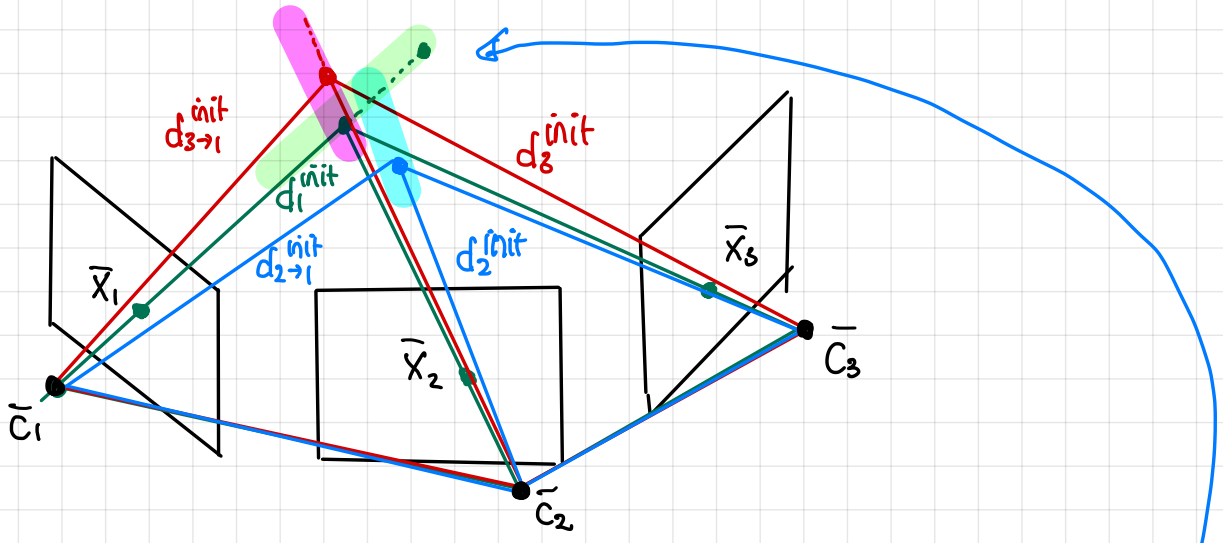


Ideally : $d_1^{init} = d_{2\to1}^{init} = d_{3\to1}^{init}$

Reality : $d_1^{init} \neq d_{2\to1}^{init} \neq d_{3\to1}^{init}$ → This is the cause of the flickering problem

To address the problem we need to make :

$d_1 \approx d_{2\to1} \approx d_{3\to1}$ → Enforce them to be as close as possible.



Basic idea : To find other disparity value other than d_t^{init} that close the gap in d_1 & $d_{2\to1}$ & $d_{3\to1}$

Candidates for d_t^{init} are : $\{d_t^{init} - N\varepsilon, \ldots, d_t^{init}, d_t^{init} + \varepsilon, \ldots, d_t^{init} + N\varepsilon\}$

N & ε are variables that need to be decided.

- Input: $\{I_t, P_t, D_t^{init}\}_{t=1}^{N}$

  Output: D_t

using the same functions as before

$$D_t = \underset{\{d_t^{init} + n\varepsilon\}_{n=-N}^{+N}}{\operatorname{argmin}} \sum_{\bar{x}} \left[\left(1 - U(\bar{x})\, f_d(\bar{x}, D_t(\bar{x}))\right) + \sum_{\bar{y} \in N\bar{x}} \lambda(\bar{x},\bar{y})\, f_p(D_t(\bar{x}), D_t(\bar{y})) \right]$$

candidates: $\{d_{t-N\varepsilon}^{init}, \ldots, d_{t-\varepsilon}^{init}, d_t^{init}, d_{t+\varepsilon}^{init}, \ldots, d_{t+N\varepsilon}^{init}\}$

Where:

$$f_d(\bar{x}, d) = \sum_{t'}^{N_t} f_c(\bar{x}, d, I_t, I_{t'})\, f_v(\bar{x}, d, D_{t'})$$

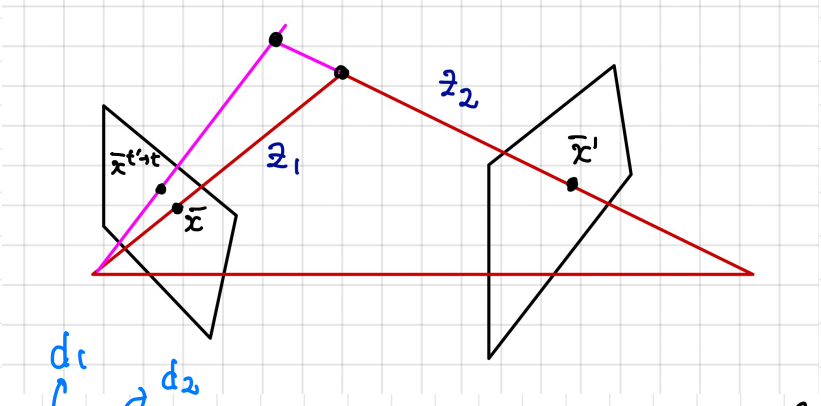
photo consistency — geometric coherence

⇓

the same as in the previous step.

Geometric Coherence:

Goal: to measure how consistent z_1 (or d_1) and z_2 (or d_2) geometrically.



d_1 d_2

$$f_v(\bar{x}, d, D_{t'}) = \exp\left(-\frac{\|\bar{x} - l_{t',t}(\bar{x}', D_{t'}(\bar{x}))\|^2}{2\sigma_d^2}\right)$$

similar to σ_c: to control how tolerant we are with the error

; σ_d = standard deviation

where: $\bar{x}' = l_{t,t'}(\bar{x}, d)$

$$\bar{x}^{t'\to t} = l_{t',t}(\bar{x}', D_{t'}(\bar{x}))$$

$$\|\bar{x} - \bar{x}^{t'\to t}\| = \|\bar{x} - l_{t',t}(\bar{x}', D_{t'}(\bar{x}))\|$$